Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012

Identifying the Semantics of Eligibility Criteria of Clinical Trials based on relevant Medical Ontologies

Anca Bucur and Jasper van Leeuwen
Healthcare Information Management
Philips Research Europe,
High Tech Campus 34
Eindhoven, the Netherlands

David Perez-Rey and Raul Alonso Calvo
Grupo de Informática Biomédica
Dept Inteligencia Artificial, Facultad de Informática,
Universidad Politécnica de Madrid, Boadilla del Monte,
Madrid, Spain

Brecht Claerhout and Kristof de Schepper
Custodix NV
Kortrijksesteenweg 214b3,
Sint-Martens-Latem, Belgium

Abstract—**An important objective of the INTEGRATE project[1] is to build tools that support the efficient execution of post-genomic multi-centric clinical trials in breast cancer, which includes the automatic assessment of the eligibility of patients for available trials. The population suited to be enrolled in a trial is described by a set of free-text eligibility criteria that are both syntactically and semantically complex. At the same time, the assessment of the eligibility of a patient for a trial requires the (machine-processable) understanding of the semantics of the eligibility criteria in order to further evaluate if the patient data available for example in the hospital EHR satisfies these criteria. This paper presents an analysis of the semantics of the clinical trial eligibility criteria based on relevant medical ontologies in the clinical research domain: SNOMED-CT, LOINC, MedDRA. We detect subsets of these widely-adopted ontologies that characterize the semantics of the eligibility criteria of trials in various clinical domains and compare these sets. Next, we evaluate the occurrence frequency of the concepts in the concrete case of breast cancer (which is our first application domain) in order to provide meaningful priorities for the task of binding/mapping these ontology concepts to the actual patient data. We further assess the effort required to extend our approach to new domains in terms of additional semantic mappings that need to be developed.**

Keywords-clinical trials, eligibility criteria, oncology, concepts, medical ontologies, SNOMED-CT, MedDRA, LOINC.

I. INTRODUCTION

The collaborative INTEGRATE project aims to support a novel research approach in oncology through the development of innovative biomedical infrastructures enabling multidisciplinary collaboration, management and large-scale sharing of multi-level data, and the development of new methodologies and of predictive multi-scale models in cancer. The INTEGRATE infrastructure will bring together heterogeneous multi-scale biomedical data generated through standard and novel technologies within post-genomic clinical trials and seamlessly link to existing research and clinical infrastructures, such as clinical trial systems, eCRFs, and hospital EHRs, in order to enable a range of innovative applications.

The project also aims to make relevant steps towards semantic interoperability. To be able to reuse previous efforts in data sharing, modeling and knowledge generation, and to access relevant external sources of data and knowledge it is beneficial to adhere whenever possible to widely-accepted standards and ontologies. The use of standards will also support wide scale adoption of our solutions.

An important objective of this project is to build tools that facilitate efficient the execution of post-genomic multi-centric clinical trials in breast cancer. A range of such tools aim to support recruitment through the automatic evaluation of the eligibility of patients for trials based on matching the characteristics of the patient population required by the trial to the patient data available for instance in the hospital EHR.

Clinical trials are key instruments in clinical research that enable the validation of research hypotheses turning them into evidence that can be applied in wide clinical care. The population suitable to be enrolled in a trial is described by a set of free-text eligibility criteria that are both syntactically and semantically complex and whose automatic evaluation in order to assess the eligibility of a patient for a set of trials is a challenging task.

As criteria describe characteristics of the eligible patient population that need to be matched against the data items that

[1] Driving excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures. http://www.fp7-integrate.eu/

are known for an individual patient, this task would be facilitated by the ability to identify the semantic entities that sufficiently describe the meaning of the criteria and by establishing links to relevant available data, stored for instance in an EHR system. Building these links (mappings) is a partially manual process and it is beneficial to be able to reuse them whenever possible across trials and systems.

This paper focuses on the analysis of the semantics of the eligibility criteria of clinical trials based on widely used medical ontologies. We identify the subsets of the ontologies that sufficiently capture the content of the eligibility criteria of trials in the clinical domain of interest which is breast cancer and compare with trials in cancers other than breast and in the cardiovascular domain.

We evaluate whether our modular approach for the selection of the sets of concepts based on the clinical domain is scalable and feasible. Selecting subsets instead of using entire ontologies facilitates the linkage of the clinical trial criteria to the actual patient records. The definition of mappings or other processing steps for entire ontologies is not feasible because of the sizes of the ontologies.

Our approach to identifying relevant subsets of ontologies relies on the annotation with ontology concepts of a large corpus of clinical trial eligibility criteria. We prioritize relevant concepts based on their frequency in the breast cancer subset and on their co-occurrence in trials in other domains.

II. The Semantic Solution and the Patient Recruitment Application

At the centre of the semantic solution that links trial descriptions to the patient data is the core dataset as in [1]: Soundly defined and agreed-upon clinical structures consisting of standard-based concepts, their relationships, quantification etc., that together sufficiently describe the clinical domain. To maximize reuse we evaluate the ability of several ontologies to capture the semantics of the criteria. It is of interest to identify the subsets of the ontologies that cover the meaning of the criteria in relevant clinical domains, the sizes of these subsets, the frequencies of concepts across trials and the overlap between subsets that describe criteria of trials in different domains. This information enables us to evaluate the effort required for the implementation of mappings, the priorities in building these mappings and the scalability of our solution with the number of trials and the extensibility to other domains.

We aim to capture the semantics of the clinical terms by standard terminology systems such as SNOMED-CT[4], MedDRA[5] and LOINC[6], which are widely used in the clinical domain. The scalability of the solution needs to be achieved by modularization, e.g. instead of aiming at inclusion of the complete SNOMED terminology we will identify a core subset that covers the chosen clinical domain and the datasets in our repositories. In the process of identifying the core dataset and the corresponding mapping tools, we need to allow for easy extension of this core dataset when the inclusion of new concepts becomes necessary (e.g. when adding new trials).

The selection of this core dataset is both clinical domain- and application-specific. Our first application area is clinical trial recruitment. To support automatic assessment of the suitability of patients for trials we need to be able to capture the semantics of the eligibility criteria and to evaluate if those are satisfied by the available patient data.

After identifying the concepts that define the semantics of the criteria we need to bind those to the information model of the system containing the patient data. As the development of these mappings is a time consuming and partially manual process it is important to minimize the effort required. Therefore, we need to evaluate the sizes of the concept sets that are relevant and the ease of handling updates (e.g. adding new clinical trials, and incorporating changes/updates in the ontologies used or in the information models of the sources) and extensions to new clinical domains. These aspects are important to assess the feasibility of our solution. In this paper we try to answer some of these questions by evaluating the semantic content of the trial eligibility criteria based on widely-used ontologies.

III. Description of the Experiment and of the Dataset

In order to analyze the semantics of eligibility criteria of clinical trials we have selected a large set of trial descriptions out of those published on ClinicalTrials.gov, a service of the U.S. National Institute of Health. We have used ClinicalTrials.gov because this site is widely used by the clinical research community and the set of trials available is both comprehensive and representative for our applications.

We selected trials from three clinical domains: breast cancer, cancer other than breast cancer, and heart and blood diseases. TABLE I. indicates the number of trials in each of the three domains. The breast cancer corpus was selected as relevant because it is the first domain for which we will implement our semantic solution and trial recruitment tools. The second corpus, clinical trials that study cancer other than breast cancer, and the third, trials that investigate heart and blood diseases, will enable us to compare the semantics of the different domains and to evaluate our modular approach and the extensibility to a new clinical domain.

TABLE I. Number of Trials in the Evaluation

Clinical domain		
Breast cancer (BC)	***Cancer other than breast (CwoBC)***	***Heart and blood diseases (HBD)***
4232	6691	12255

We extracted the eligibility criteria from these sets of trials and used a state of the art annotator to identify the ontology concepts present in these criteria. The annotator is available at BioPortal [2] and is developed by the National Center for Biomedical Ontology. The BioPortal annotation results include information such as the concept name, concept identifier and the UMLS[3] semantic type of the concept.

[2] http://bioportal.bioontology.org/

[3] http://www.nlm.nih.gov/research/umls/

The annotator allows the user to select out of a library of over 300 biomedical ontologies those that are relevant and should be used in the annotations. We have selected SNOMED-CT, MedDRA and LOINC.

We extracted and analyzed the sets of ontology concepts that were found to link to items from the eligibility criteria of our selected collection of clinical trials and compared the result for the three clinical domains selected and the three medical ontologies. The results of this analysis are described in the next section.

IV. Evaluation Results

The first step in the analysis was to identify the sizes of the sets of concepts that describe the semantics of a domain and how much of the entire ontology they represent. Our modular approach to semantic linkage would not work if a large part of those ontologies is relevant for the trial criteria, for instance because implementing semantic mappings for a large ontology such as SNOMED-CT (over 311 000 concepts in 2011) requires a huge effort and would not be feasible for our application.

Next, we have compared the subsets of concepts among the different domains and for the three ontologies to identify overlaps and extensions. This enables us to estimate the effort of implementing our solution for the initial domain of breast cancer and the ease of extending this solution to new domains.

Another aspect of interest is to compare trials in each domain and assess how similar the semantics of distinct trials are. A large degree of similarity (which would be expected) means that once implementing our solution for a sufficiently large set of trials, adding new trials requires little effort. It is also relevant to prioritize the concepts that are occurring most often.

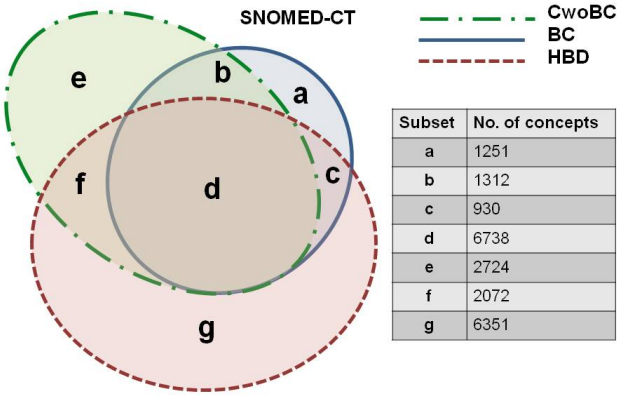
Finally, we investigated the most frequent semantic types that correspond to the concepts identified in the criteria. This additional information is relevant as it enables us to classify concepts with similar content or from similar sources.

A. *Subsets of concepts*

The figures below compare the sets of relevant concepts for the three clinical domains and the three ontologies that we selected. In all cases the largest set is the one that is the overlap among the three domains, therefore concepts that are domain independent. The breast cancer corpus has a small subset that is specific for this disease (marked with "a" in the figure) and also relatively small subsets that constitute overlaps with each of the other domains. This makes our modular approach very feasible as a large amount of the concepts used in the semantic solution for breast cancer will be also relevant for other diseases.

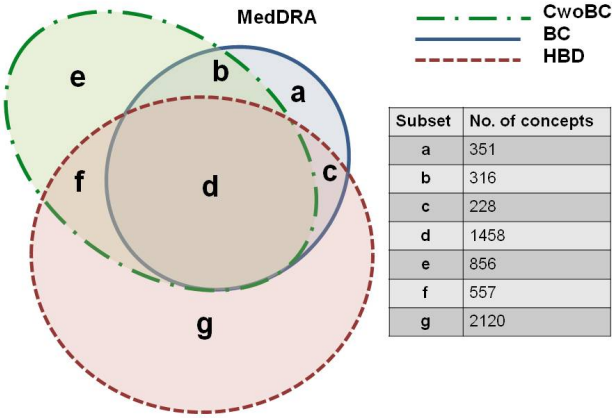
Extensions to new domains are also manageable as the additional sets of concepts are relatively small, even for completely different domains (e.g. extending from BC to HBD). This is especially the case for LOINC, where the module covering the concepts that are specific for HBD is half the size of the overlap with BC and CwoBC.

Also in absolute numbers, the sizes of the sets of concepts that capture the semantics of our domains are reasonable and support the implementation of our semantic solution and of the trial recruitment applications that will rely on it.



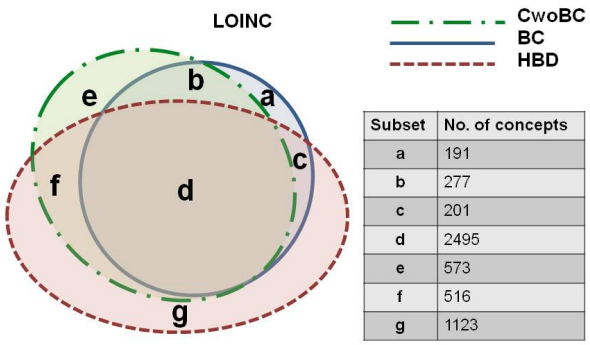
Subset	No. of concepts
a	1251
b	1312
c	930
d	6738
e	2724
f	2072
g	6351

Figure 1. Sets of SNOMED-CT concepts for breast cancer (BC), cancer other than breast cancer (CwoBC) and heart and blood disease (HBD)



Subset	No. of concepts
a	351
b	316
c	228
d	1458
e	856
f	557
g	2120

Figure 2. Sets of MedDRA concepts for breast cancer (BC), cancer other than breast cancer (CwoBC) and heart and blood disease (HBD



Subset	No. of concepts
a	191
b	277
c	201
d	2495
e	573
f	516
g	1123

Figure 3. Sets of LOINC concepts for breast cancer (BC), cancer other than breast cancer (CwoBC) and heart and blood disease (HBD)

Error! Reference source not found.indicates the ratio of the three selected ontologies that cover the semantics of our domains. It shows that a small percentage of the ontologies are sufficient to capture the content of the trial descriptions in the specific domains. For instance, in the case of breast cancer trials 3.2 % of SNOMED CT, 3.3% of MedDRA and 5.4% of LOINC were used in the annotation of our datasets.

TABLE II. RATIO OF THE ONTOLOGIES THAT CAPTURE THE SEMANTICS OF THE THREE SETS OF CLINICAL TRIALS

	SNOMED-CT	***MedDRA***	***LOINC***
BC	0.032	0.033	0.054
CwoBC	0.041	0.045	0.066
HBD	0.051	0.062	0.074

B. Reoccurrence of concepts across trials

In this section we investigate the semantic similarity among trials. Intuitively we expected that trials will have a large ratio of criteria that are similar, but that new trials do introduce new concepts. This is confirmed by Figure 4. that depicts for the three corpuses of trials and the three ontologies the distribution of concepts across trials. Only the top most frequent concepts are depicted and each concept is counted once per trial. In all cases there is a relatively small group of concepts that occur in a large number of trials and there is another group of concepts that are rare or unique for specific trials. To further illustrate this, TABLE III. provides for each selected clinical domain and ontology the average number of trials in which a concept occurs, the average number of trials for the top 100 most frequent concepts, and the average number of trials for the top 500 most frequent concepts.

To have an additional reference, we have also counted the number of ontology concepts that occur per trial. In the case of breast cancer we have concluded that there are on average 199 SNOMED-CT concepts per trial, 27 MedDRA concepts and 108 LOINC concepts.

These facts enable us to prioritize the implementation of semantic mappings starting with the concepts that occur often and demonstrate that the effort of adding new trials is low: Updates will be required, but the additional concepts that need to be mapped to relevant data are few.

TABLE IV. and TABLE V. include examples of very frequent concepts of SNOMED-CT and MedDRA that occur in the BC dataset.

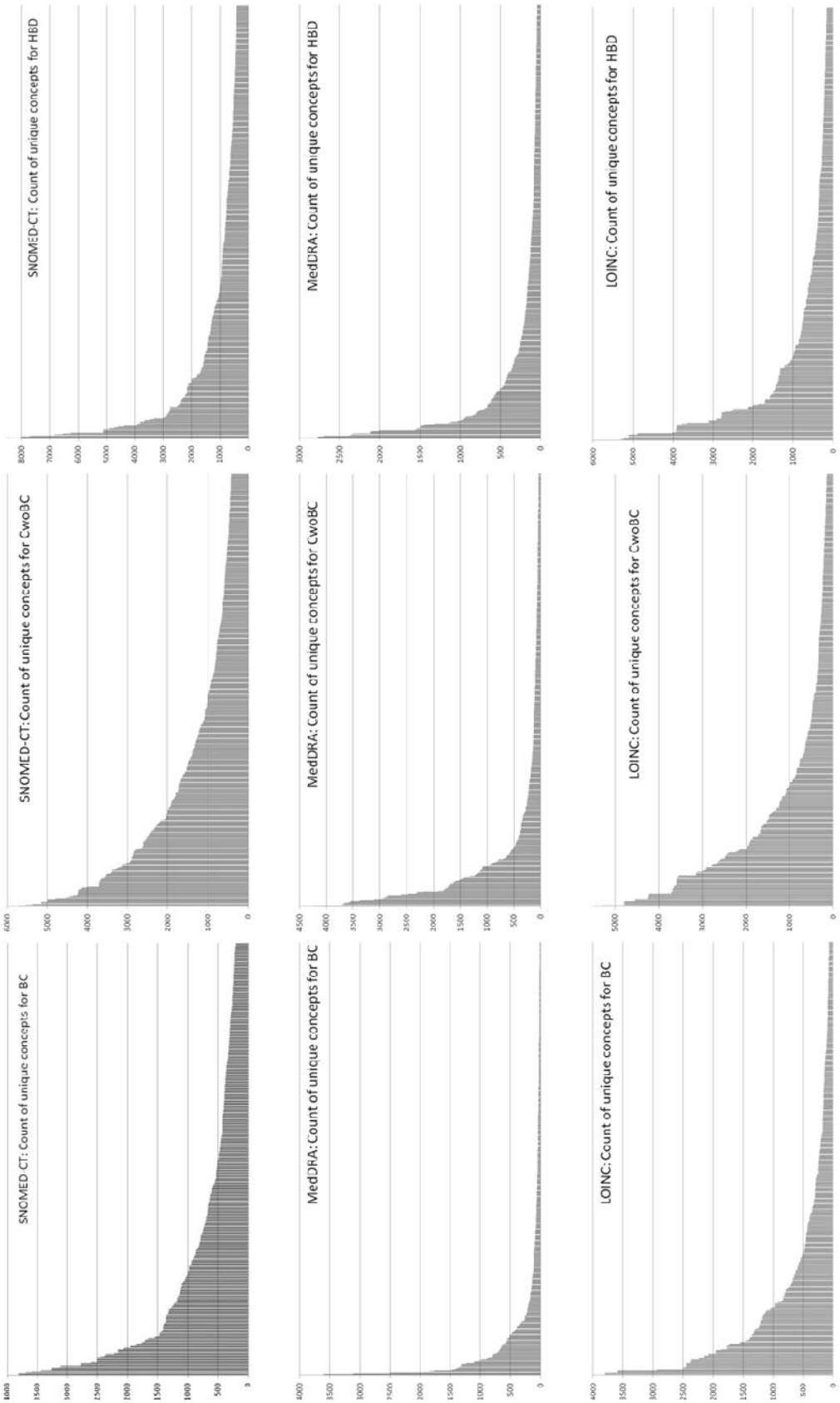


Figure 4. The reoccurrence of concepts across trials. The number of BC, CwoBC and respectively HBD trials (y-axis) that include the top 500 most frequently occuring concepts (x-axis) out of SNOMED-CT, MedDRA and LOINC. Concepts were counted once per trial.

TABLE III. STATISTICS OF THE REOCCURENCE OF ONTOLOGY CONCEPTS IN BREAST CANCER TRIALS

Reoccurrence of ontology concepts for BC (4232 trials)			
Statistics	***SNOMED-CT***	***MedDRA***	***LOINC***
Average number of trials for top 100 concepts	1835.57	632.61	1597.45
Average number of trials for top 500 concepts	756.63	163.01	510.11
Average number of trials for all unique concepts	51.865	36.98	89.31
Number of distinct concepts	10231	2353	3164
Concepts occuring in a single trial	3159	815	671

TABLE IV. FREQUENT SNOMED-CT CONCEPTS IN BREAST CANCER TRIALS (ALL OCCURENCES COUNTED)

Concept name	***Concept code***	***Number of occurrences***
Disease	64572001	13261
Neoplasm, malignant (primary)	86049000	11241
Entire breast	181131000	10254
Breast structure	76752008	10246
Therapeutic procedure	103733002	9958
Therapy	276239002	9956
Malignant neoplastic disease	363346000	9764
History of	392521001	8170
Study	224699009	7823
Malignant tumor of breast	254837009	5544
Antineoplastic chemotherapy regimen	69960004	5544
Drug therapy	182831000	5336

TABLE V. FREQUENT MEDDRA CONCEPTS IN BREAST CANCER TRIALS (ALL OCCURENCES COUNTED)

Concept name	***Concept code***	***Number of occurrences***
Cancer	10007050	9737
Breast cancer	10006187	6133
Chemotherapy	10061758	5320
Metastatic	10027474	3689
Carcinoma	10007284	2759
Surgery	10042609	2739
Radiotherapy	10037794	2612
Pregnant	10036586	2148
Metastases	10027476	2109
Creatinine	10011358	1956

C. Semantic types

In this section we evaluate the UMLS semantic types of the concepts in our datasets as they can provide additional information about the semantics of the criteria and identify concepts that are similar. We compare the frequency in the sets of concepts of several semantic types that are relevant for our application domain. We have annotated with semantic types all the concepts identified in all three ontologies.

The tables below depict the most frequent semantic types for the three corpuses: BC, CwoBC and HBD. We can observe that there are little differences among the clinical domains in the hierarchy of semantic types to which most of the concepts belong.

TABLE VI. RATIO OF THE MOST FREQUENT UMLS SEMANTIC TYPES FOR THE BC DATASET (RELATIVE TO TOTAL NUMBER OF CONCEPTS)

Semantic type	*Ratio*
Neoplastic Process	0.055
Qualitative Concept	0.041
Therapeutic or Preventive Procedure	0.039
Body Part, Organ, or Organ Component	0.031
Laboratory Procedure	0.023
Finding	0.018
Pharmacologic Substance	0.012
Diagnostic Procedure	0.010
Quantitative Concept	0.006
Spatial Concept	0.005

TABLE VII. RATIO OF THE MOST FREQUENT UMLS SEMANTIC TYPES FOR THE CWOBC DATASET (RELATIVE TO TOTAL NUMBER OF CONCEPTS)

Semantic type	*Ratio*
Qualitative Concept	0.080
Therapeutic or Preventive Procedure	0.076
Neoplastic Process	0.065
Laboratory Procedure	0.049
Body Part, Organ, or Organ Component	0.041
Finding	0.032
Diagnostic Procedure	0.022
Pharmacologic Substance	0.022
Quantitative Concept	0.012
Spatial Concept	0.010

TABLE VIII. RATIO OF THE MOST FREQUENT UMLS SEMANTIC TYPES FOR THE HBD DATASET (RELATIVE TO TOTAL NUMBER OF CONCEPTS)

Semantic type	*Ratio*
Therapeutic or Preventive Procedure	0.044
Qualitative Concept	0.041
Body Part, Organ, or Organ Component	0.038
Finding	0.036
Laboratory Procedure	0.029
Pharmacologic Substance	0.020
Diagnostic Procedure	0.018
Quantitative Concept	0.015
Spatial Concept	0.011
Sign or Symptom	0.010

V. RELATED WORK

In this paper we focus on clinical trials in breast cancer and present an analysis of the semantics of the eligibility criteria of trials based on widely-adopted medical ontologies: SNOMED-CT[4], MedDRA[5] and LOINC[6].

SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) is a clinical vocabulary focused on accurately recording health care encounters and the associated electronic health information exchange. Although SNOMED-CT is sometimes criticized, it has a significant uptake in clinical practice, such as its use in HL7[7] messaging. MedDRA focuses on the regulatory process of drug development and is a medical vocabulary that is used by regulatory bodies and the regulated biopharmaceutical industry for data entry, retrieval, evaluation and display. MedDRA is used in clinical trials for reporting adverse events. LOINC (Logical Observation Identifiers Names and Codes) has the purpose to facilitate the exchange and pooling of results for clinical care, outcomes management, and research. LOINC provides universal identifiers for laboratory and other clinical observations and it is a preferred code set for HL7 for laboratory test names in transactions between health care facilities, laboratories, laboratory testing devices, and public health authorities.

With respect to the selection of domain-specific parts of ontologies, in [2] the subset of UMLS that is relevant to describe breast cancer treatment was identified in order to facilitate the development of clinical decision support systems. While the general idea is comparable, the purpose and the method were different. As background knowledge the concepts from medical guidelines were used, considered at the decision points of selecting a suitable treatment for a patient. The guideline concepts were manually mapped to the SNOMED-CT concepts and the subset obtained was automatically expanded via the ontology hierarchy and the UMLS semantic network.

A significant body of research has focused on the general problem of formalization of eligibility criteria and on

[4] http://www.ihtsdo.org/SNOMED CT/
[5] http://www.meddramsso.com
[6] http://loinc.org/
[7] http://www.hl7.org/

recruitment for clinical trials, including (semi)-automatic trial matching. In [3] an extensive overview of existing solutions and approaches is provided. In previous work [4] we have analyzed the eligibility criteria of clinical trials and we have identified relevant syntactic patterns that occur in trial criteria. These patterns are modifiers that provide the context of the criterion and while they do not express the semantics of the criterion and cannot be linked to actual patient data for evaluation of eligibility, they provide the context of the criterion. We have evaluated the coverage of these patterns for a large set of eligibility criteria and their expressivity.

In [5] we evaluate the approach to eligibility criteria formalization, with the focus on pattern detection. The feasibility is influenced by the number of patterns, the algorithm of patterns detection, and the restrictiveness and the variety of synonym forms covered by the regular expressions. We conclude that the algorithm of patterns detection has sufficient precision and recall to support, in most cases, the generation of correct corresponding queries and to determine patient eligibility. The broad range of defined patterns allows automatic interpretation of even complex eligibility criteria.

In [6] an analysis has been carried out to estimate the coverage provided by SNOMED-CT for clinical research concepts that represented by the items present on case report forms (CRFs). The authors also evaluated the semantic nature of those concepts relevant to post-coordination methods. The dataset included a total of 17 CRFs developed by rheumatologists conducting several longitudinal, observational studies in the clinical domain of vasculitis. From the CRFs a total set of 616 (unique) items were identified. Each unique data item was classified as either a clinical finding or procedure. The items were coded by the presence and nature of SNOMED CT coverage and manually classified into semantic types by 2 coders.

In [1] we introduce a scalable, modular and pragmatic approach to achieving semantic interoperability. We believe that interoperability in healthcare can be achieved gradually on specific domains and by making use whenever possible of existing standards. This is also the approach that we take in the INTEGRATE project for a well-defined clinical domain which is clinical trials in breast cancer. As presented in this paper, we identify those modules of ontologies that are relevant in this domain and in our semantic solution we will implement mappings for those specific concepts. This facilitates efficient further extensions to other domains of relevance and easy reuse of tools. A gradual approach to interoperability is well supported in literature. In [7] it is stated that "regardless of the type of vision one may develop, semantic interoperability is not a phenomenon to be expected over night". The group of experts conclude that semantic interoperability in healthcare requires a large number of changes at both the technical and the use case level, and that even in that vision no full semantic interoperability or a complete harmonization of either EHR models or terminologies can be expected.

VI. CONCLUSIONS

This paper has focused on the evaluation of the semantics of the eligibility criteria of clinical trials in terms of identifying relevant ontology concepts that express the content of the criteria. In the context of developing applications supporting efficient execution of clinical trials it is essential to assess whether our modular semantic linkage approach is applicable to this domain. This requires to decide whether the semantics of the eligibility criteria can be captured by widely-used medical ontologies and to estimate the effort required to semantically link the eligibility criteria to the relevant patient data (for example preserved in an EHR) to enable the decision of whether the patient satisfies the criteria.

The ontologies selected were SNOMED-CT, MedDRA and LOINC which are widely used in the clinical domain and when used by our solutions could support scalability and adoption. We have identified the relevant subsets of these ontologies that capture the semantics of the eligibility criteria of clinical trials in selected clinical domains.

Another important question we have answered is of extendibility. Our main focus in the INTEGRATE project is breast cancer, but we aim to design solutions that can be extended and applied to other clinical domains. Therefore we evaluated and compared the sets of concepts that capture the semantics of different clinical domains: breast cancer, cancer other than breast cancer, and heart and blood disease.

The analysis of the concepts that are specific to a domain or occur across various domains let us modularize the sets of concepts that are relevant for a particular group of trials. We identified the subset of concepts that exclusively occur in eligibility criteria related to one of the three domains, those that are shared among trials in various clinical domains.

We have relied on the annotation of a large collection of clinical trials using the NCBO's BioPortal annotator. Our findings indicate that relatively small subsets (in terms of number of concepts) of the ontologies are required to capture the semantics of the eligibility criteria. It was also shown that the semantic overlap among clinical domains is very large for all ontologies considered, therefore once developed for a particular domain a large part of the mappings can be reused when extending the solution to a new domain. The additional sets of concepts that are specific to those domains are relatively small and the implementation of the new mappings is feasible.

The frequency of the concepts, their reoccurrence across various trials and their uniqueness for particular types of trials informs the selection of the concept sets that cover the meaning of the criteria. These statistics guide the process of linking the concepts to the data items in the patient records by building the necessary mappings.

We have concluded that the reuse of concepts across trials is very significant, with a relatively small number of concepts that occur in many trials. Therefore, they can be prioritized in the implementation of mappings. We can capture a large part of the semantics of the trials with a relatively small number of concepts that sufficiently describe the content of the eligibility criteria.

The infrequent concepts that are specific to single trials are also manageable and it is most efficient for the implementation of the semantic solution to only add those when a trial containing them is entered into the system. The long tail of the graph indicates that the sets of concepts identified will not be complete and will grow with new trials, but the high overlap across trials makes the effort of handling updates for new trials low.

We have also evaluated the UMLS semantic types of the concepts as these can provide additional hints about the semantics of the criteria and can be used in the semantic solution to reason about the criteria at a higher level of abstraction. We compared the frequency in the sets of concepts of several semantic types that are relevant for our application domain.

In our future work we will assess the coverage of our domain of interest by the sets of concepts, identify and address limitations (post-coordinated terms in SNOMED, criteria not covered by any concept, etc.), and develop the necessary mappings between the clinical trial criteria and our HL7 RIM-based patient data model.

ACKNOWLEDGEMENTS

This work has been partially funded by the European Commission through the INTEGRATE project (FP7-ICT-2009-6-270253).

REFERENCES

[1] R. Vdovjak, B. Claerhout, and A. Bucur, "Bridging the Gap between Clinical Research and Care - Approaches to Semantic Interoperability, Security & Privacy", in HEALTHINF 2012, pp. 281-286.

[2] K. Milian, Z. Aleksovski, R. Vdovjak, A. ten Teije, and F. van Harmelen, "Identifying disease-centric subdomains in very large medical ontologies", in Proceedings of the AIME'09 workshop on Knowledge Representation for Healthcare (KR4HC09). 2009; pp. 41-50.

[3] C. Weng, S.W. Tu, I. Sim, and R. Richesson, "Formal representation of Eligibility Criteria: A Literature Review", in Journal of Biomedical Informatics, 2010.

[4] K. Milian, A. ten Teije, A. Bucur, and F. van Harmelen, "Patterns of clinical trial eligibility criteria", in Proceedings of the AIME'11 workshop on Knowledge Representation for Healthcare (KR4HC11), lecture notes AI, 2011.

[5] K. Milian, A. Bucur, and A. ten Teije, "Formalization of clinical trial eligibility criteria: Evaluation of a pattern-based approach" in Proceedings of the BIBM '12, 2012

[6] R.L. Richesson, J.E. Andrews, and J.P. Krischer, "Use of SNOMED CT to Represent Clinical Research Data: A Semantic Characterization of Data Items on Case Report Forms in Vasculitis Research", in JAMIA 13(5), 2006, pp. 536-546.

[7] V.N. Stroetmann et al., "Semantic Interoperability for Better Health and Safer Healthcare", SemanticHEALTH Report, 2009, http://ec.europa.eu/information_society/activities/health/docs/publications/2009/2009semantic-health-report.pdf.